

Association genetics of oleoresin flow in loblolly pine: discovering genes and predicting phenotype for improved resistance to bark beetles and bioenergy potential

Jared W. Westbrook¹, Marcio F. R. Resende Jr¹, Patricio Munoz¹, Alejandro R. Walker^{1,2}, Jill L. Wegrzyn³, C. Dana Nelson⁴, David B. Neale³, Matias Kirst^{1,2}, Dudley A. Huber², Salvador A. Gezan², Gary F. Peter^{1,2} and John M. Davis^{1,2}

¹Forest Genomics Laboratory, Genetics Institute, University of Florida, 1376 Mowry Rd, Rm 320, Gainesville, FL 32611, USA; ²School of Forest Resources and Conservation, University of Florida, PO Box 110410, Gainesville, FL 32611, USA; ³Department of Plant Sciences, University of California at Davis, Mail Stop 4, Davis, CA 95616, USA; ⁴Southern Institute of Forest Genetics, USDA Forest Service, Southern Research Station, 23332 Success Rd, Saucier, MS 39574, USA

Author for correspondence:

John M. Davis

Tel: +1 352 846 0879

Email: jmdavis@ufl.edu

Received: 18 January 2013

Accepted: 15 February 2013

New Phytologist (2013) 199: 89–100

doi: 10.1111/nph.12240

Key words: association genetics, genomic selection, genotype \times environment interactions, herbivory defense, *Pinus*, terpenoid.

Summary

- Rapidly enhancing oleoresin production in conifer stems through genomic selection and genetic engineering may increase resistance to bark beetles and terpenoid yield for liquid bio-fuels.
- We integrated association genetic and genomic prediction analyses of oleoresin flow ($\text{g } 24 \text{ h}^{-1}$) using 4854 single nucleotide polymorphisms (SNPs) in expressed genes within a pedigreed population of loblolly pine (*Pinus taeda*) that was clonally replicated at three sites in the southeastern United States.
- Additive genetic variation in oleoresin flow ($h^2 \approx 0.12\text{--}0.30$) was strongly correlated between years in which precipitation varied ($r_a \approx 0.95$), while the genetic correlation between sites declined from 0.8 to 0.37 with increasing differences in soil and climate among sites. A total of 231 SNPs were significantly associated with oleoresin flow, of which 81% were specific to individual sites. SNPs in sequences similar to ethylene signaling proteins, ABC transporters, and diterpenoid hydroxylases were associated with oleoresin flow across sites.
- Despite this complex genetic architecture, we developed a genomic prediction model to accelerate breeding for enhanced oleoresin flow that is robust to environmental variation. Results imply that breeding could increase oleoresin flow 1.5- to 2.4-fold in one generation.

Introduction

In the last decade, outbreaks of bark beetles in coniferous forests of North America have caused unprecedented tree mortality and economic losses (Nowak *et al.*, 2008; van Mantgem *et al.*, 2009; Waring *et al.*, 2009), converting forests that were previously atmospheric carbon sinks into carbon sources (Kurz *et al.*, 2008). Native species of bark beetle rapidly kill healthy trees by aggregating on their hosts, boring into the stem, and vectoring pathogenic fungi that are tolerant of conifer defenses (Paine *et al.*, 1997; Wang *et al.*, 2013). Climate change is thought to have exacerbated tree mortality from bark beetle infestations by increasing the number of beetle generations yr^{-1} , expanding the range of beetles and their associated pathogens, and by weakening host defenses (Raffa *et al.*, 2008; Bentz *et al.*, 2010).

Increasing oleoresin production in conifer stems through breeding and biotechnology may enhance baseline resistance to bark beetles in managed plantations (Phillips & Croteau, 1999). Oleoresin is a viscous mixture of terpenoids stored under positive

pressure within resin canals in the stems of conifers (Trapp & Croteau, 2001). Oleoresin that flows from the stem upon wounding obstructs beetle entry and inhibits germination and growth of pathogenic fungi (Franceschi *et al.*, 2005; Kopper *et al.*, 2005). Among pines, survival after a bark beetle infestation is positively correlated with the rate of oleoresin flow (Strom *et al.*, 2002) and the number of resin canals within the stem (Kane & Kolb, 2010). However, increased stem oleoresin production alone may not be sufficient to protect conifer stands from severe bark beetle population eruptions, where healthy trees with the greatest oleoresin flow can become the preferred hosts of beetles (Boone *et al.*, 2011).

The potential to utilize terpenoids in liquid biofuels is an additional incentive to genetically enhance oleoresin production in conifer stems. Whereas the energy content of bioethanol is only 70% of that of gasoline (Peralta-Yahya & Keasling, 2010), advanced biofuels derived from terpenoids have similar energy content to gasoline and diesel, with densities and hygroscopicities that are amenable to blending with fossil fuels (Harvey *et al.*,

2010; Peralta-Yahya *et al.*, 2011). Conifers require fewer inputs of fertilizer and herbicide than annual food crops even under intensive management, and the net energy balance of producing cellulosic ethanol from conifer wood compares favorably with ethanol derived from maize starch (Evans & Cohen, 2009).

The capacity to genetically enhance terpenoid production in conifer stems is largely untapped. Under normal growing conditions, pines accumulate oleoresin to 1–5% of their stem mass, but stem oleoresin contents of 20% have been observed after treatment with chemical elicitors of resinosis (Stubbs *et al.*, 1984; Wolter & Zinkel, 1984). Although it remains to be demonstrated whether these high terpenoid concentrations can be achieved through genetics, previous studies indicate that variation in oleoresin flow is heritable and positively correlated with growth (Squillace & Bengtson, 1961; Roberds *et al.*, 2003; Romanelli & Sebbenn, 2004).

Developing a detailed knowledge of how allelic variation in conifer breeding populations relates to phenotypic variation in oleoresin flow can accelerate selective breeding for enhanced oleoresin production. Conifer breeding has traditionally relied on phenotypic characterization of the breeding population near the harvest age to infer genetic merit for selection, requiring ≥ 20 yr to complete one breeding cycle (White & Carson, 2004). Genomic selection, or the prediction of breeding values from the summed effects of a panel of genetic markers in linkage disequilibrium with alleles controlling a trait (Meuwissen *et al.*, 2001), circumvents the need to phenotype the breeding population for each generation. When combined with top-grafting to induce early seed development, genomic selection could reduce the breeding cycle of *Pinus taeda* from 12–20 yr to 4–7 yr (Resende *et al.*, 2012a).

Genetic engineering is an alternative strategy to rapidly increase oleoresin production in conifer stems. This approach may include overexpressing and increasing the catalytic efficiency of terpenoid biosynthetic enzymes (Aharoni *et al.*, 2006; Leonard *et al.*, 2010), down-regulating competing pathways (e.g. lignin biosynthesis), and reprogramming stem development to favor resin canal formation (Zulak & Bohlmann, 2010). While many genes in the terpenoid biosynthetic pathway have been cloned in conifers (Kim *et al.*, 2008, 2009; Schmidt & Gershenzon, 2008; Hamberger *et al.*, 2011; Keeling *et al.*, 2011), the genes involved in the regulation of terpenoid synthesis (e.g. transcriptional and post-translational regulation), the development of resin canals, and the transport of oleoresin into resin canals have yet to be characterized. Coordinated up-regulation of the terpenoid biosynthetic pathway and resin canal development may be achieved through transgenic manipulation of regulatory genes, as has generally been suggested for the metabolic engineering of plant defense pathways (Jirschitzka *et al.*, 2012).

The genomic selection and biotechnological approaches to increase oleoresin production in conifers are potentially complicated by genotype \times environment ($G \times E$) interactions. Substantial $G \times E$ in oleoresin flow was observed among families of *Pinus sylvestris* and *Pinus elliotti* planted at different sites (Bridgen, 1980; Romanelli & Sebbenn, 2004). For traits in which $G \times E$ is prevalent, the effects of alleles may depend on

the environment (Gillespie & Turelli, 1989), which reduces the prediction accuracy of genomic selection models (Resende *et al.*, 2012a), and leads to uncertainty about the performance of transgenic varieties in diverse field environments (Zeller *et al.*, 2010).

We measured oleoresin flow in a pedigreed population of loblolly pine that was clonally replicated at three sites in the southeastern United States and estimated heritability, site \times genotype interactions, and genetic correlations with growth. By comparing genetic correlations in oleoresin flow across years vs between sites, we assessed the relative effect of weather vs climate and soils, respectively, on the genetic control of oleoresin flow. We then used an association genetic approach to discover genes underlying additive genetic variation in oleoresin flow and compared their allelic effects among sites. We specifically tested for associations with single nucleotide polymorphisms (SNPs) in genes with potential roles in terpenoid biosynthesis. Finally, we used cross-validation to assess how accurately the significantly associated loci could predict additive genetic variation within and across sites.

Materials and Methods

Oleoresin flow phenotyping in the CCLONES population

Oleoresin flow and tree growth were measured in the loblolly pine (*Pinus taeda* L.) CCLONES (Comparing Clonal Lines ON Experimental Sites) series I population. The CCLONES population was generated from a circular mating design of 54 parents from Florida, the Atlantic coastal plain, and lower gulf regions of the southeastern United States. Each parent was crossed with two to six other parents to generate 71 full-sib families with 14–23 progeny per family (Baltunis *et al.*, 2007; Munoz *et al.*, 2013). Clones of these progeny were sampled for oleoresin flow at three US field sites (Cuthbert, GA, Nassau, FL, and Palatka, FL), which varied in climate and in soil properties in the summer of 2010 when the trees were in their seventh growing season (Table 1). The Nassau site was also sampled in the sixth growing season to estimate correlations among repeated measures of oleoresin flow between years that varied in precipitation (Table 1). Within each site, oleoresin flow was sampled from three intensively managed replicates of clonal genotypes (Baltunis *et al.*, 2007). The position of clonal genotypes (ramets) within replicates was randomized to incomplete blocks of 10–14 trees according to a resolvable alpha-lattice design (Patterson & Williams, 1976).

To sample oleoresin flow, a single circular wound, 1.27 cm in diameter, was made through the bark and phloem at breast height with an arch punch (Strom *et al.*, 2002; Roberds *et al.*, 2003). Immediately after wounding, taps with 15 ml tubes were affixed to the stem over the wound site and oleoresin samples were collected after 24 h. Oleoresin samples were weighed twice – within 1 wk of collection (wet mass) and after the samples were lyophilized for 3 d (dry mass) – to detect any potential bias resulting from occasional rainwater contamination and phenotyping errors (Supporting Information, Fig. S1a). The heritability was slightly higher for the dry mass measurements (Fig. S1b) and the

Table 1 Description of the *Pinus taeda* CCLONES sites and oleoresin flow sampling

	Cuthbert, GA	Nassau, FL	Palatka, FL
Location and date			
Latitude	31°47'32"N	30°46'30"N	29°38'24"N
Longitude	84°41'32"W	81°53'30"W	81°49'27"W
Sampling dates	26–29 July 2010	20–23 July 2009 19–22 July 2010	16–19 August 2010
Climate and weather			
Annual minimum temperature	–9.5 to –12.5°C	–6.7 to –9.4°C	–3.9 to –6.7°C
Average temperature, 1 July–31 July in year of sampling	Year 7: 28.7°C	Year 6: 27.5°C Year 7: 28.9°C	Year 7: 28.4°C
Precipitation, 1 February–31 July in year of sampling	Year 7: 547 mm	Year 6: 918 mm Year 7: 537 mm	Year 6: 649 mm
Long-term average precipitation, 1 February–31 July	737 mm	707 mm	635 mm
Soil properties			
Soil drainage	Well drained	Somewhat poorly drained	Poorly to very poorly drained
Soil taxonomy	Fine-loamy, kaolinitic, thermic Rhodic Kandiuult	Fine, mixed, thermic Typic, Albaqualf	Sandy, siliceous, hyperthermic Ultic Alaquod
Stand properties – year 6			
Trees ha ^{–1}	1653	1700	1725
Basal area (m ² ha ^{–1})	130.4	129.3	91.4
Oleoresin sampling design			
N_{ramets}	Year 7: 2437	Year 6: 1829 Year 7: 2575	Year 7: 2586
N_{clones}	Year 7: 847	Year 6: 891 Year 7: 908	Year 7: 913
$N_{\text{genotyped}}$	Year 7: 809	Year 6: 836 Year 7: 850	Year 7: 842

Precipitation and temperature data were compiled from cocorahs.org. Weather stations were within 20 km of the CCLONES sites. Annual minimum temperatures were derived from the USDA plant hardiness zones map (www.usna.usda.gov/Hardzone/). N_{ramets} , number of ramets with oleoresin dry mass, total tree height (ht), and diameter at breast height (dbh) measurements (up to three ramets sampled per clone within a site); N_{clones} , number of clonal genotypes sampled; $N_{\text{genotyped}}$, number of phenotyped clones with single nucleotide polymorphism (SNP) genotype information.

additive genetic correlation between oleoresin wet mass and dry mass was 0.99, indicating that lyophilization did not change the genetic rankings of the clones. Therefore, oleoresin dry mass (in g) was used in quantitative genetic analyses. Diameter at breast height (dbh) and total tree height (ht) were measured in cm in the sixth growing season from the trees that were sampled for oleoresin.

SNP genotyping

A total of 7535 SNPs, discovered in expressed genes from loblolly pine of diverse origin (Eckert *et al.*, 2010), were genotyped with the Illumina Infinium platform (Illumina, San Diego, CA, USA) in 926 clones. A subset of 4854 SNP loci that were polymorphic in CCLONES was utilized in association analyses and genomic prediction modeling. Mean and median missingness of SNP genotypes among loci were 3.81 and 1.3%, respectively.

Quantitative genetic analysis

Oleoresin dry mass, dbh, and ht data were analyzed in ASReml v.3 (Gilmour *et al.*, 2009) with the following multivariate mixed models:

$$\mathbf{y} = \mathbf{1}\boldsymbol{\mu} + \mathbf{X}_r\mathbf{r} + \mathbf{Z}_b\mathbf{b} + \mathbf{Z}_c\mathbf{a} + \mathbf{Z}_f\mathbf{f} + \mathbf{Z}_e\mathbf{c} + \mathbf{e}$$

$$\mathbf{y} = \mathbf{1}\boldsymbol{\mu} + \mathbf{X}_s\mathbf{s} + \mathbf{X}_r\mathbf{r} + \mathbf{Z}_b\mathbf{b} + \mathbf{Z}_c\mathbf{a} + \mathbf{Z}_f\mathbf{f} + \mathbf{Z}_e\mathbf{c} + \mathbf{Z}_{sa}\mathbf{sa} + \mathbf{Z}_f\mathbf{sf} + \mathbf{e}$$

Data from individual sites were analyzed with the first model to estimate site-specific heritabilities and correlations among traits. The data from all sites were combined and analyzed with the second model to partition site \times genotype interactions from genetic effects. A Box–Cox power transformation (Box & Cox, 1964) was applied to the oleoresin dry mass data (hereafter abbreviated as ‘tdm’ for transformed dry mass) to normalize the model residuals (Methods S1). For both the individual site and across site models, \mathbf{y} is a $3 \times N_{\text{ramets}}$ (Table 1) matrix of observed phenotypes of tdm, dbh, and ht, and \mathbf{X} and \mathbf{Z} are incidence matrices associating phenotypic observations with fixed and random factors, respectively. The effects of fixed factors, which included $\boldsymbol{\mu}$ (the overall trait means), \mathbf{s} (site effects), and \mathbf{r} (the effects of replicate within site), were assessed with approximate F -tests (Kenward & Roger, 1997). Random effects were modeled with unstructured variance-covariance matrices of tdm, dbh, and ht and assumed to be multivariate normal with mean 0. Random factors (followed by their phenotypic variance components)

included **b** (incomplete block within site and replicate ($\sigma_b^2 \mathbf{I}$)), **a** (additive genetic effect of clone ($\sigma_a^2 \mathbf{A}_{obs}$)), **f** (nonadditive genetic effect of family ($\sigma_f^2 \mathbf{I}$)), **c** (nonadditive genetic effect of clone ($\sigma_c^2 \mathbf{I}$)), **sa** (site \times additive genetic ($\sigma_{sa}^2 \mathbf{I}$)), **sf** (site \times family ($\sigma_{sf}^2 \mathbf{I}$)), and **e** (error ($\sigma_e^2 \mathbf{I}$)). Additive genetic variance (σ_a^2) and estimated breeding values (EBVs) for clones were calculated with the observed genomic relationship matrix (\mathbf{A}_{obs}), which models identity-by-descent (IBD) as a result of pedigree as well as Mendelian segregation within families (Visscher *et al.*, 2006). The observed relationship matrix was reconstructed from 4854 polymorphic SNP loci by modifying the method of Yang *et al.* (2010) such that Mendelian variance in marker-estimated IBD coefficients was estimated from the deviations from IBD expected from the pedigree. The individual site model was modified to estimate correlations between years at Nassau (\mathbf{y} – tdm from years 6 and 7) and correlations between sites (\mathbf{y} – tdm from Cuthbert, Nassau, and Palatka in year 7). Table S1 contains the formulas for estimating heritabilities and correlations. The significance of correlations was assessed by likelihood ratio tests (Methods S2).

Association genetic analysis

Analysis to detect significant associations between individual SNP loci and additive genetic variation in tdm within and across sites was carried out in two stages. First, all 4854 SNPs were ranked according to the magnitude of their effect on additive genetic variation in tdm. A total of 400 SNPs, which included the most highly ranked loci and 16 SNPs in sequences similar to terpenoid biosynthetic enzymes, were selected for the second stage where missing SNP genotypes were imputed and the significance of SNP effects was assessed in a multiple regression framework with the Bayesian Association with Missing Data (BAMD) module for R (Gopal *et al.*, 2011; Li *et al.*, 2012). The purpose of SNP preselection was to reduce multicollinearity and computational times in the BAMD analyses (Li *et al.*, 2012).

We tested three SNP preselection methods (additive genetic variance reduction, ridge regression, and Bayes C π). Additive genetic variance reduction tests each SNP individually for the proportion of additive genetic variance explained, while ridge regression and Bayes C π estimate SNP effects simultaneously with multiple regression. Ridge regression shrinks SNP effects toward zero and assumes that the effects are normally distributed, whereas Bayes C π estimates the proportion of SNPs with zero effect, and then assumes that effects of the remaining SNPs are multivariate Student's *t* distributed (Habier *et al.*, 2011; Methods S3). Among preselection methods, SNPs selected with Bayes C π that were significant in BAMD analysis most accurately predicted additive genetic variation in tdm in cross-validation (Fig. S2). Therefore, only the association results using SNPs selected with Bayes C π are reported.

Association analyses were conducted in BAMD with the model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\gamma + \varepsilon$$

The response variable \mathbf{y} comprised the best linear unbiased predictions (BLUPs) of estimated breeding values (EBVs) of tdm

divided by their reliabilities (hereafter referred to as deregressed EBVs). Dividing EBVs by their reliability ($r_i^2 = 1 - [\text{SEP}_i / \sqrt{\sigma_a^2}]^2$, where SEP_i is the standard error of prediction for clone $i = 1 \dots N_{\text{clones}}$, $0 < r_i^2 < 1$) corrected for variable shrinkage of BLUPs as a result of imbalance in the number of phenotypic observations per clone and in the amount of information available from relatives (Garrick *et al.*, 2009). No subpopulation structure was detected in CCLONES (Methods S4); therefore, it was unnecessary to include an additional effect of subpopulation to control for spurious associations arising from population structure (Pritchard *et al.*, 2000).

The SNP effects were modeled by the term $\mathbf{X}\gamma$. The incidence matrix \mathbf{X} contained the SNP genotypes of the 400 preselected loci. BAMD iteratively imputed missing genotypes with a Gibbs sampler over 50 000 generations based on correlations among the observed SNP values, phenotypic observations, and the additive genetic relationships (Li *et al.*, 2012). After each imputation, SNP effects were estimated assuming a normal distribution of effect sizes. An SNP locus was deemed to have a significant effect on \mathbf{y} if the 95% confidence interval of estimated effects over the last 40 000 generations did not intersect zero. Significant loci were ranked according to the magnitude of their standardized effects $\gamma_{\text{std}} = \frac{1}{2} (\text{CI}_{\text{upper}} + \text{CI}_{\text{lower}}) / \sigma_a$, where CI_{upper} and CI_{lower} are the upper and lower bounds of the 95% confidence interval of effect size, and σ_a is the additive genetic standard deviation of the trait.

Population structure from the pedigree and uncertainty in the estimation of breeding values was modeled in the error $\varepsilon \sim N(0, \sigma_e^2 \mathbf{R})$ where $\mathbf{R} = \mathbf{A}_{\text{exp}} + \mathbf{W}$. The expected relationship matrix (\mathbf{A}_{exp}) modeled correlation among the residuals as a result of pedigree structure and the diagonal matrix \mathbf{W} weighted deregressed EBVs in inverse proportion to the BLUP variances after factoring out error variance as a result of nongenetic factors (Garrick *et al.*, 2009).

Functional annotation and mapping of significant loci

Expressed sequence tags that contained the significant SNPs were assessed for similarity to genes of known function with Blast2GO (Conesa *et al.*, 2005). A subset of the significant SNPs from each site were located on a *P. taeda* genetic map from Eckert *et al.* (2010), which included 1495 of the 4854 SNPs used in association analysis.

Prediction of breeding values from genetic markers within and among sites

Prediction of breeding values from marker genotypes within sites was implemented in R (R Core Team, 2012) with 10-fold cross-validation as described in Resende *et al.* (2012a). Briefly, clones present within each site were randomly divided into 10 groups. Marker effects were estimated with ridge regression of marker genotypes on deregressed EBVs from nine-tenths of the population. Parental effects were removed from the deregressed EBVs to correct for gametic phase disequilibrium of unlinked markers arising from pedigree structure (Garrick *et al.*, 2009). Breeding

values from the remaining one-tenth of the population were predicted by multiplying the incidence matrix of marker genotypes by the vector of estimated marker effects. This procedure was repeated 10 times to predict breeding values for all clones present within a site. To predict breeding values among sites, the genotypes of markers significantly associated with tdm were regressed on EBVs from the site where they were significant. The estimated marker effects were then used to predict the breeding values from other sites. The accuracy of prediction within and among sites was assessed with the Pearson correlation between deregressed EBVs and SNP-estimated breeding values ($r_{EBV, GEBV}$).

Results

Phenotypic variation in constitutive oleoresin flow and cumulative growth

The untransformed oleoresin dry mass distributions were positively skewed at all sites (Fig. 1a), but were approximately normally distributed after transformation. Transformed oleoresin dry mass (tdm) differed significantly among sites ($P < 0.001$ for all pairwise site contrasts, Table S2), but site rankings for tdm from year 7 (Nassau > Palatka > Cuthbert) did not correspond to site rankings for dbh (Cuthbert > Nassau > Palatka) (Fig. 1b). At Nassau, where oleoresin flow was measured in consecutive years, mean tdm was significantly greater in year 7, during a dry summer, relative to wet growing season in year 6 (Tables 1, S2).

Genetic and site \times genotype effects on constitutive oleoresin flow

Within sites, additive genetic variance (σ_a^2) accounted for 11.8–30.3% of the phenotypic variance in tdm, and nonadditive genetic variance ($\sigma_f^2 + \sigma_c^2$) explained an additional 8.4–17.1% of the phenotypic variance. Across sites, phenotypic variance in tdm consisted of 11.8% additive genetic variance, 9.4% nonadditive genetic variance, and 6.9% site \times genotype variance ($\sigma_{sa}^2 + \sigma_{sf}^2$).

The heritabilities and site \times genotype effects on tdm were comparable to those of dbh and ht (Fig. 2).

Repeated measurements of tdm at the Nassau site were moderately correlated between years 6 and 7 ($r_p = 0.41$, Fig. 3a), in which precipitation varied (Table 1); however, strong additive ($r_a = 0.95$) and total genetic ($r_g = 0.90$) correlations were observed across years (Fig. 3b). Stochastic variation in tdm between years was attributable mainly to environmental effects (Fig. 3c, Table 1). By contrast, the additive genetic correlation of tdm between sites declined from 0.8 to 0.37 (Fig. 4) with increasing differences in soil and climate among sites (Table 1).

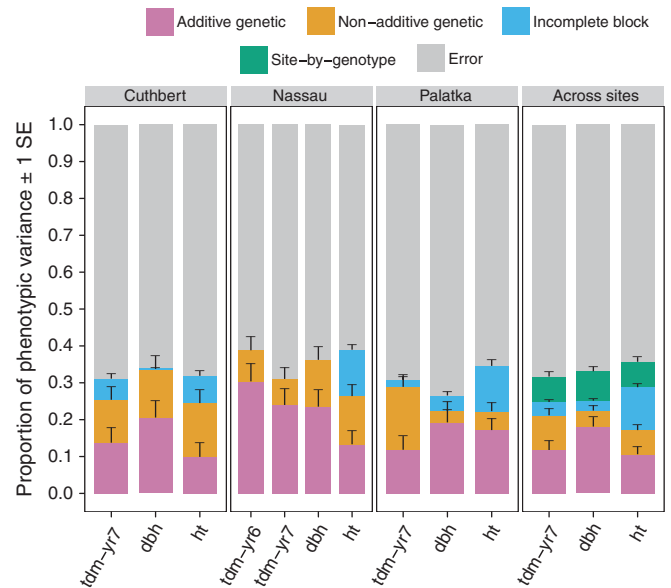


Fig. 2 Proportion of phenotypic variance in transformed oleoresin dry mass (tdm), diameter at breast height (dbh) and total tree height (ht) comprising additive genetic (σ_a^2), nonadditive genetic ($\sigma_f^2 + \sigma_c^2$), site \times genotype ($\sigma_{sa}^2 + \sigma_{sf}^2$), incomplete block (σ_b^2), and environmental error (σ_e^2) variance components estimated in the *Pinus taeda* CCLONES population. Refer to Table S1 for equations to estimate variance proportions.

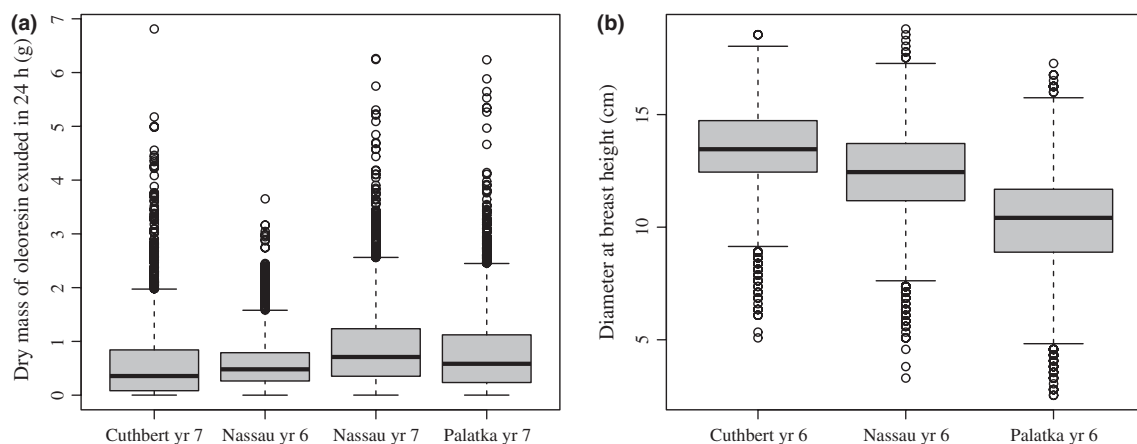


Fig. 1 Boxplot distributions of oleoresin dry mass by site and year (a) and diameter at breast height by site (b) measured at age 6 yr in the *Pinus taeda* CCLONES population. Outlying data points, denoted by circles, fall outside the interval defined by the interquartile range (IQR) $\pm 1.5 \times$ IQR.

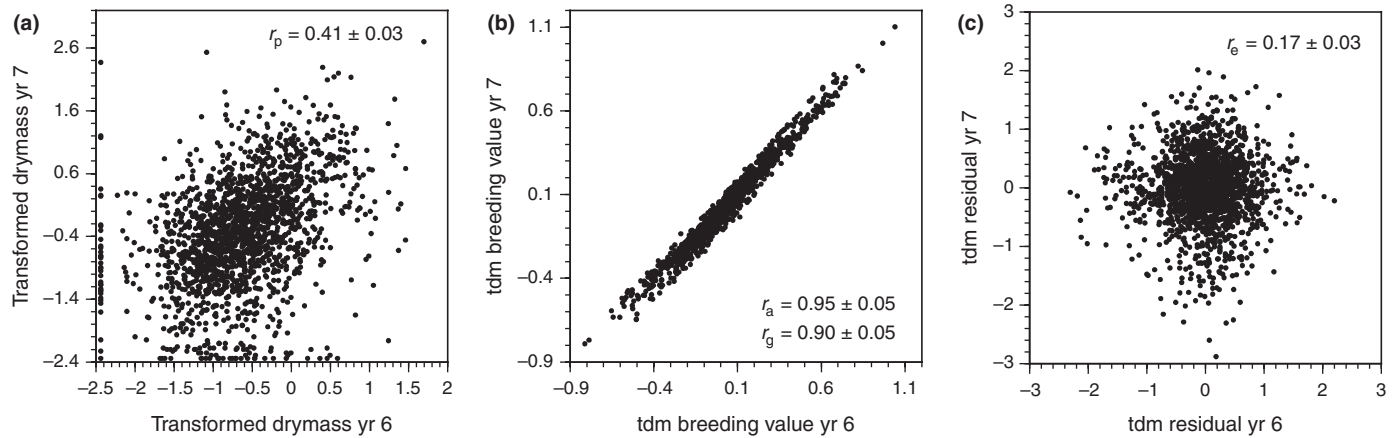


Fig. 3 Phenotypic, genetic, and environmental correlations ± 1 SE of transformed constitutive oleoresin dry mass (tdm) measured from the same ramets in consecutive years (ages 6 and 7 yr) in the *Pinus taeda* CCLONES population at Nassau, FL, USA. (a) Phenotypic correlation (r_p) adjusted for the fixed effect of replication; (b) genetic correlation between years (r_a , additive genetic; r_g , total genetic) visualized by plotting tdm breeding values; and (c) environmental correlation (r_e) visualized by plotting the model residuals. All correlations were estimated according to equations in Table S1 and were significant ($P < 0.05$) according to the likelihood ratio test.

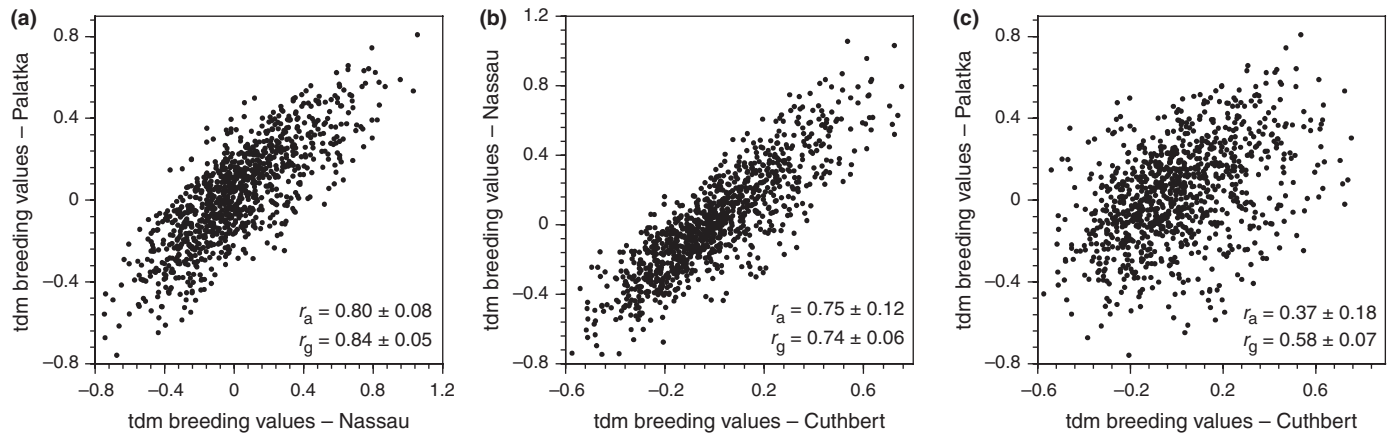


Fig. 4 Genetic correlations (r_a , additive genetic; r_g , total genetic) ± 1 SE between transformed oleoresin dry mass (tdm) estimated from *Pinus taeda* CCLONES genotypes that were clonally replicated at different sites. (a–c) Panels are ordered from sites that are closest geographically to sites that are most distant. Breeding values of tdm are plotted to visualize the genetic correlations between sites, which were estimated from the equations in Table S1. All correlations were significant ($P < 0.05$) according to the likelihood ratio test.

Correlations between oleoresin flow and cumulative growth

Genetic correlations between oleoresin dry mass and cumulative tree growth (i.e. ht and dbh in an even-aged stand) were positive at Nassau and Palatka, but were not significant at Cuthbert and in the combined site analysis. Site \times genotype and environmental effects on tdm and growth were also positively correlated. Generally, correlations between tdm and tree size were weaker than correlations between dbh and ht (Table 2).

Associations between SNPs and additive genetic variation in oleoresin flow

Between 41 and 65 SNPs were associated with tdm at individual sites, and 73 SNPs were associated with tdm in the across-site analysis (Table 3). While few significant tdm associations were

shared between sites (two to six common associations), greater numbers of common associations were detected between years at Nassau (10 common associations) and between individual sites and the across-site analysis (eight to 20 common associations). The additive genetic effects of individual SNP loci were small, ranging from 0.05 to 0.15 additive genetic standard deviations (Table S3). No changes in the sign of the significant SNP effects were detected between sites, and correlations in the magnitude of common SNP effects varied from -0.375 to 0.9 among site pairs (Table 3). A total of 231 SNP loci were associated with tdm among all analyses (Table S3), and 132 of these loci were located on the genetic map from Eckert *et al.* (2010). Mapped SNPs that were significantly associated with tdm were scattered throughout the genome (Fig. 5).

Many significant tdm associations were unique to specific sites (Table 3; Fig. 5). Site-specific associations may be attributed to unbalanced representation of clonal genotypes at different sites,

Table 2 Correlations (± 1 SE) between transformed oleoresin dry mass and cumulative growth in the *Pinus taeda* CCLONES population

	r_a	r_g	r_e	r_p	r_{sg}
Cuthbert, GA					
tdm vs dbh	0.26 ± 0.14	0.17 ± 0.09	$0.20 \pm 0.03^*$	$0.19 \pm 0.03^*$	–
tdm vs ht	0.16 ± 0.25	0.18 ± 0.14	$0.19 \pm 0.03^*$	$0.19 \pm 0.03^*$	–
ht vs dbh	$0.73 \pm 0.11^*$	$0.72 \pm 0.04^*$	$0.65 \pm 0.02^*$	$0.67 \pm 0.02^*$	–
Nassau, FL					
tdm vs dbh	$0.43 \pm 0.13^*$	$0.39 \pm 0.07^*$	$0.27 \pm 0.03^*$	$0.31 \pm 0.03^*$	–
tdm vs ht	$0.55 \pm 0.15^*$	$0.34 \pm 0.08^*$	$0.19 \pm 0.02^*$	$0.24 \pm 0.03^*$	–
ht vs dbh	$0.67 \pm 0.10^*$	$0.71 \pm 0.04^*$	$0.63 \pm 0.02^*$	$0.65 \pm 0.02^*$	–
Palatka, FL					
tdm vs dbh	0.19 ± 0.18	$0.26 \pm 0.08^*$	$0.32 \pm 0.02^*$	$0.32 \pm 0.02^*$	–
tdm vs ht	$0.39 \pm 0.19^*$	$0.39 \pm 0.07^*$	$0.31 \pm 0.02^*$	$0.31 \pm 0.02^*$	–
ht vs dbh	$0.86 \pm 0.04^*$	$0.82 \pm 0.03^*$	$0.83 \pm 0.01^*$	$0.83 \pm 0.01^*$	–
Across sites					
tdm vs dbh	0.15 ± 0.13	$0.22 \pm 0.07^*$	–	–	$0.40 \pm 0.11^*$
tdm vs ht	0.27 ± 0.15	$0.31 \pm 0.07^*$	–	–	$0.36 \pm 0.12^*$
ht vs dbh	$0.75 \pm 0.06^*$	$0.75 \pm 0.03^*$	–	–	$0.72 \pm 0.06^*$

An asterisk (*) next to a correlation indicates that the correlation was significant ($P < 0.05$) by likelihood ratio test (Methods S2). Environmental and phenotypic correlations were not estimated across sites because environmental deviations were assumed to be independent among sites. tdm, Transformed oleoresin dry mass; ht, total tree height; dbh, diameter at breast height; r_a , additive genetic correlation; r_g , total genetic correlation; r_e , environmental correlation (within sites); r_p , phenotypic correlation (within sites); r_{sg} , site \times genotype correlation (across sites).

Table 3 Single nucleotide polymorphisms (SNPs) associated with transformed oleoresin dry mass within and between sites in the *Pinus taeda* CCLONES population

	Across sites	Cuthbert	Nassau, year 6	Nassau, year 7	Palatka
Across sites	73	0.038	0.900	0.478	0.551
Cuthbert	12	65	NA	–0.375	0.630
Nassau, year 6	8	2	41	0.354	0.169
Nassau, year 7	20	5	10	65	0.120
Palatka	15	5	4	6	55

Diagonal elements are the number of SNPs significantly associated with transformed oleoresin dry mass (tdm) at each site. Elements below the diagonal are the number of loci that were significant in pairs of association analyses. Elements above the diagonal are the correlations among the magnitude (i.e. absolute values) of the effects of SNPs that were significant between analyses.

statistical significance of different SNP loci linked to the same quantitative trait locus (QTL) at different sites, and site \times QTL interactions. To control for imbalance in clonal genotypes among sites, the association analyses of tdm from individual sites were repeated with 722 clones that were present at all three sites. Furthermore, a common set of 157 mapped SNPs, where adjacent loci were > 10 cM apart (12.5 cM average distance between adjacent loci), were utilized in the association analysis of each site to control for linkage and SNP preselection. After implementing these controls, 81% of the tdm associations were site-specific (Table S4), supporting the hypothesis that site-specific associations are attributable to site \times QTL interactions.

Within sites, the significant SNPs predicted additive genetic variation in tdm more accurately than randomly selected SNPs

(Fig. 6). Maximum prediction accuracy was reached before the stepwise inclusion of all SNPs that were associated with tdm, and was greater than the accuracies obtained from the estimated effects of all 4854 polymorphic SNPs. Furthermore, SNPs associated with tdm in year 6 were significant predictors of tdm breeding values from year 7 at Nassau (Fig. 6b).

The accuracy by which significant SNPs from one site predicted additive genetic variation from another site (Fig. 7) was proportional to the additive genetic correlation (r_a) between sites (Fig. 4). Cuthbert and Palatka, with an additive genetic correlation of 0.37, had the lowest between-site prediction accuracies ($r_{EBV, GEBV} = 0.33$ –0.36). By contrast, Nassau and Palatka, with an additive genetic correlation of 0.80, had the greatest between-site prediction accuracies ($r_{EBV, GEBV} = 0.42$ –0.43) and the prediction accuracy between years at Nassau ($r_a = 0.95$; Fig. 3) varied from 0.4 to 0.5. Notably, the SNPs that were significant in the across-site association analysis (Table S3), in which site \times genotype effects were partitioned from additive genetic variation (Fig. 2), predicted breeding values from single sites with accuracies ($r_{EBV, GEBV} = 0.42$ –0.56) similar to within-site prediction accuracies ($r_{EBV, GEBV} = 0.51$ –0.62).

Associations with SNPs in sequences similar to terpene biosynthetic genes

We tested for tdm associations with SNPs in sequences similar to four of seven steps of the DXP pathway, three geranyl-geranyl pyrophosphate synthases (GGPPS), one terpene synthase, and seven cytochrome P450s (Table S5). The DXP pathway synthesizes the five-carbon isoprenoid precursors of mono- and diterpenoids (Rodriguez-Concepcion & Boronat, 2002); GGPPS condenses isoprenoids into the 20-carbon skeletons of diterpenoids (Schmidt & Gershenzon, 2008); terpene synthases are a

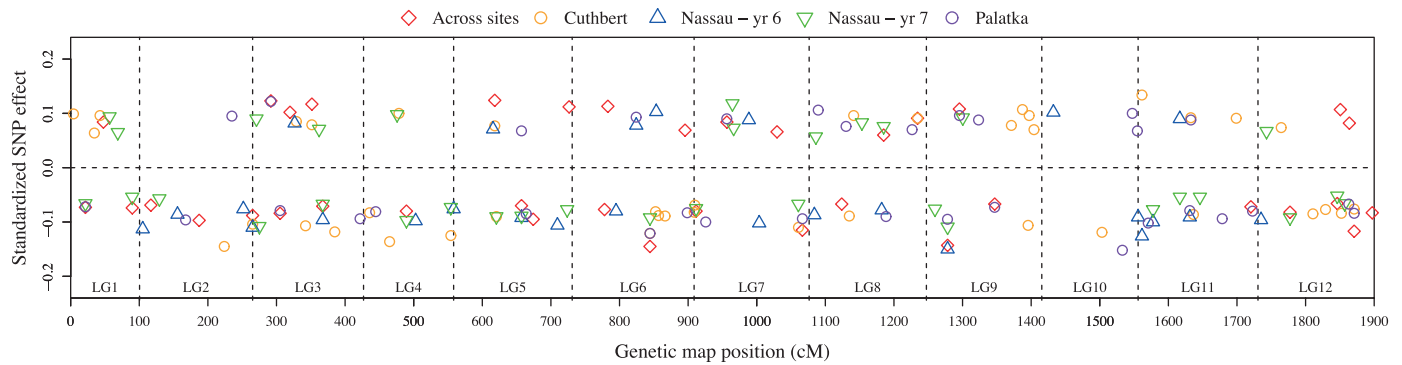


Fig. 5 Genetic map positions of single nucleotide polymorphisms (SNPs) significantly associated with transformed oleoresin dry mass and their effects by site. Of the 231 loci that were significantly associated with transformed oleoresin dry mass (tdm) among all analyses, 132 were mapped on the *Pinus taeda* linkage map from Eckert *et al.* (2010). SNP effects are in units of additive genetic standard deviations. Vertical dashed lines demarcate linkage groups (LG1–LG12).

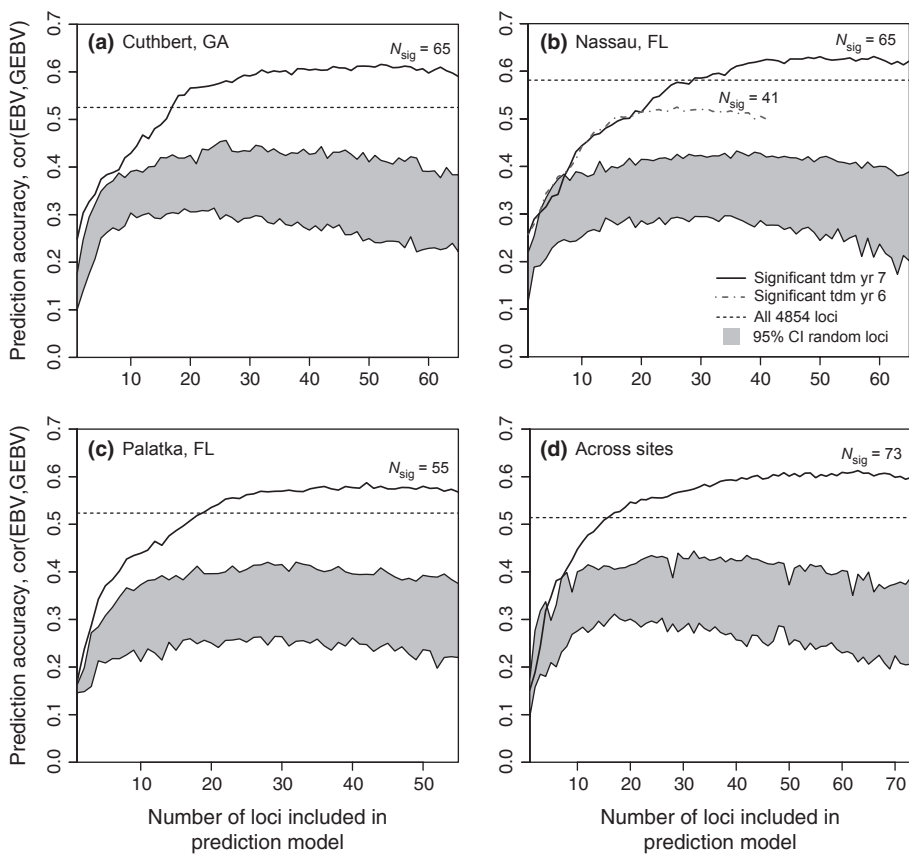


Fig. 6 Accuracies by which the summed effects of single nucleotide polymorphisms (SNPs) associated with transformed oleoresin dry mass (tdm) predicted tdm breeding values from year 7 within individual sites (a–c) and across sites (d). Prediction accuracies, measured as the Pearson correlation between genomic estimated breeding values (GEBV) and estimated breeding values from quantitative genetic analyses (EBV), were computed with 10-fold cross-validation. The prediction accuracies of SNPs associated with tdm in year 7 were compared with prediction accuracies of SNPs associated with tdm in year 6 (c), accuracies obtained from fitting all 4854 polymorphic SNP loci, and accuracies from randomly sampled loci. The 95% confidence intervals (CIs) of prediction accuracy from random loci were constructed by randomly sampling SNP subsets with replacement 1000 times. The maximum values on the x-axes are numbers of loci significantly associated with tdm within sites (N_{sig}). SNP loci were added to the prediction model in decreasing order of their average effect on tdm among 10 random partitions of the *Pinus taeda* CCLONES population.

large gene family in conifers that synthesize diverse cyclic terpenoids (Bohlmann *et al.*, 1998; Keeling *et al.*, 2011); and cytochrome P450s (CYP450) catalyze the final oxidation steps in diterpenoid resin acid synthesis (Ro *et al.*, 2005; Hamberger *et al.*, 2011). Only SNPs in two CYP450s were significantly associated with tdm. One significant CYP450 (EST contig 0_14468; Table S3) was similar to 3-epi-6-deoxocathasterone 23-monooxygenase, a gene involved in brassinosteroid biosynthesis (Kim *et al.*, 2005). The other CYP450 (contig 2_9017) was similar to taxane 13- α -hydroxylase, a gene involved in the synthesis of the diterpenoid taxol (Jennewein *et al.*, 2000).

Discussion

Phenotypic variation in oleoresin flow is heritable

Our results confirm that variation in oleoresin flow is heritable in loblolly pine and can therefore be increased through selective breeding. Although additive genetic variation accounted for only 12–30% of phenotypic variation within sites and 12% of the variation across sites (Fig. 2), we predict that oleoresin flow could be increased 1.5- to 2.4-fold in one generation by crossing clones in the 90th to 99th percentile of the additive genetic distribution

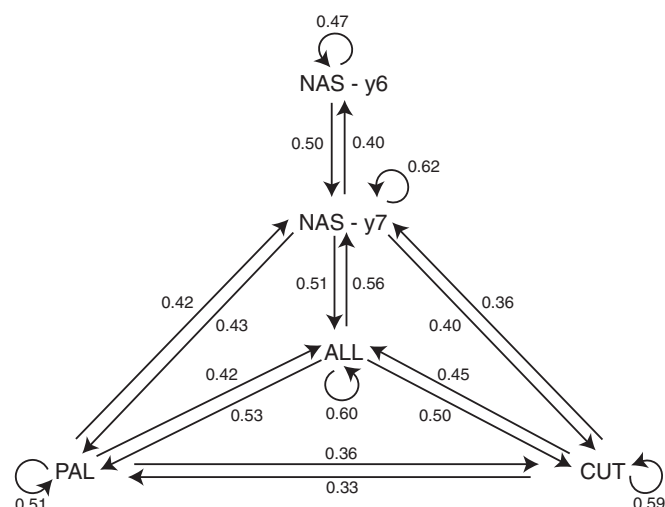


Fig. 7 Prediction accuracies of single nucleotide polymorphism (SNP) loci significantly associated with transformed oleoresin dry mass (tdm) across sites and years in the *Pinus taeda* CCLONES population. The effects of significant loci were estimated with ridge regression on deregressed estimated breeding values (EBVs) from the site where the loci were significant. These locus effects were then used to predict the deregressed EBVs from another site. Prediction accuracies of SNP loci within sites where the loci were significant (circular arrows) were estimated with 10-fold cross-validation (Fig. 6). CUT, Cuthbert, GA; NAS, Nassau, FL; PAL, Palatka, FL; ALL, across-site analysis.

(Methods S5, Table S6). These large predicted gains from selection are the result of a strong positive skew in the phenotypic distribution of oleoresin flow (Fig. 1; Roberds *et al.*, 2003).

The within-site broad-sense heritability estimates reported here ($H^2 = 0.25\text{--}0.38$; Fig. 2) are lower than previous heritability estimates of oleoresin flow from a loblolly pine progeny test grown on one site in Florida ($H^2 = 0.66\text{--}0.69$; Roberds *et al.*, 2003). Roberds *et al.* (2003) averaged two observations of oleoresin mass per tree before estimating heritability, which reduced the within-individual component of variation. By contrast, we sampled each tree once per time point, but sampled up to three clonal replicates of each genotype per site. Environmental variation within and between cloned genotypes may have reduced our heritability estimates in comparison to Roberds *et al.* (2003), but clonal replication yielded more precise estimates of breeding values and a way to examine genotype \times environment interactions (Gezan *et al.*, 2006; Baltunis *et al.*, 2007).

Effects of weather vs climate and soils on the genetic control of oleoresin flow

Within the Nassau site, average oleoresin flow was greatest in the seventh growing season (Fig. 1a; Table S2), which was relatively dry compared with the previous year (Table 1). This result supports the hypothesis that moderate water deficit increases oleoresin flow by limiting growth more than photosynthesis (Lorio, 1986; Lorio & Sommers, 1986). The strong genetic correlation in tdm between years at Nassau (Fig. 3b) indicates that oleoresin flow in loblolly pine genotypes of diverse origin responds similarly to variation in weather. By contrast,

substantial site \times genotype interactions in oleoresin flow were observed between Cuthbert and Palatka, the sites that were most distinct in climate and soils (Fig. 4c, Table 1). Together, these trends indicate that soil and climatic variation are stronger drivers of genotype \times environment interactions in oleoresin flow than year-to-year variation in weather.

Genetic correlations between oleoresin flow and growth among sites

Estimates of the genetic correlation between oleoresin flow and growth varied from positive to nonsignificant among sites (Table 2), indicating that selective breeding for enhanced oleoresin production would not negatively impact growth. We hypothesize that variation in the genetic correlation between oleoresin flow and growth can be attributed, in part, to differences in tree–tree competition among sites. The weakest oleoresin flow–growth correlations (Table 2) and the lowest average oleoresin flow rates (Fig. 1a; Table S2) were observed at Cuthbert, the site with the largest trees (Fig. 1b) and highest basal area (Table 1). Previous studies have indicated that competition reduces allocation to herbivory defense (Moreno *et al.*, 2009; Aspinwall *et al.*, 2011), especially among shade-intolerant species such as loblolly pine (Calder *et al.*, 2011). Conversely, thinning increased oleoresin flow and bark beetle resistance in *Pinus ponderosa* (Wallin *et al.*, 2008). Thus, the results of this study and others (Nowak *et al.*, 2008 and references therein) suggest the importance of managing competition to fully realize the genetic capacity for oleoresin flow in managed pine plantations.

Environmental variation interacts with the genetic architecture of oleoresin flow

Association genetic analysis of transformed oleoresin dry mass yielded over 200 significant associations within and across sites (Tables 3, S3), indicating that oleoresin flow is a quantitative trait (Fig. 5). Furthermore, 81% of the significant associations were site-specific (Table S4), implying that the effects of alleles underlying quantitative genetic variation in oleoresin flow depend on the environmental context. The molecular mechanisms hypothesized to mediate genotype \times environment interactions in plants include environmental regulation of gene expression, epigenetic modification, and post-translational modification (Nicotra *et al.*, 2010). Numerous regulatory genes were associated with oleoresin flow (Table S3), including transcription factors (regulators of gene expression), genes involved in histone modification and the processing of microRNAs (regulators of epigenetic modifications), and protein kinases and phosphatases (regulators of post-translational modification), indicating that any or all of these processes might contribute to site \times genotype interactions in oleoresin flow.

Prospects for genomic selection for enhanced oleoresin flow

Genomic selection has the potential to accelerate breeding for enhanced oleoresin production by alleviating the need to phenotype the breeding population in each generation (Meuwissen

et al., 2001; Grattapaglia & Resende, 2011; Resende *et al.*, 2012a,b). Prediction models that included only the significantly associated SNPs predicted additive genetic variation in oleoresin flow with greater accuracy than models with either randomly selected SNPs or all polymorphic loci (Fig. 6). This result indicates that our association genetic pipeline efficiently selected markers linked to causative polymorphisms and suggests that SNP preselection can increase accuracy in genomic prediction modeling by reducing model overparameterization (Schulz-Streeck *et al.*, 2011; Resende *et al.*, 2012b). Despite the prevalence of site \times genotype effects (Figs 2, 4) and site-specific associations (Tables 3, S4), SNPs that were significantly associated with breeding values from the across-site quantitative genetic analysis predicted additive genetic variation at single sites with accuracies comparable to within-site prediction accuracies (Fig. 7). Thus, to develop genomic selection models that accurately predict genetic variation in diverse environments, marker effects should be estimated on breeding values in which genotype \times environment interactions have been partitioned from genetic effects.

Association analysis to discover candidate genes for transgenic manipulation

The association genetic results reported here (Table S3) are a preliminary step towards targeting candidate genes for over-expression or silencing to validate function in oleoresin production. We highlight a few genes associated with oleoresin flow that are intriguing within the context of previous research.

Ethylene synthesis and signaling Ethylene induces oleoresin synthesis and the differentiation of resin canals in members of the Pinaceae (Hudgins & Franceschi, 2004; Schmidt *et al.*, 2011), including loblolly pine (Telewski *et al.*, 1983; Stubbs *et al.*, 1984). One SNP in an ACC synthase, a rate-limiting enzyme that catalyzes the first step of ethylene synthesis, was significant across sites (contig 0_17633). Furthermore, SNPs in ETHYLENE-INSENSITIVE 2 (contig 0_14532), which is involved in the early detection of ethylene within the cell (Alonso *et al.*, 1999), and an APETALA2 domain transcription factor (contig 0_3648) that functions downstream of EIN2 (Ogawa *et al.*, 2007) were also associated with tdm (Table S3).

Oleoresin transport The mechanism(s) by which oleoresin is transported from living cells in the stem to the extracellular space of resin canals in conifers is unknown (Zulak & Bohlmann, 2010). An ATP-binding cassette (ABC) transporter (*NpABCI*) mediates the export of a diterpene (sclareol) from *Nicotiana plumbaginifolia* cells (Jasinski *et al.*, 2001); therefore, it is plausible that ABC transporters are also involved in the transport of conifer oleoresin. Variation in tdm was significantly associated with two putative ABC transporters (UMN_2415 and 0_8686).

Associations detected in an independent population Eckert *et al.* (2012) conducted association genetic analyses of stem diterpenoids in a loblolly pine population of unrelated individuals that is independent of the CCLONES population studied here.

SNPs in two genes (contigs 0_9534 and UMN_596) that were associated with stem abietic acid content in Eckert *et al.* (2012) were also associated with oleoresin flow in CCLONES. Contig 0_9534 is a putative acid phosphatase, while the function of contig UMN_596 is unknown.

Conclusion

Both genomic selection and genetic engineering to enhance oleoresin production in conifer stems could benefit from an improved understanding of the underlying genetic architecture. Although heritable variation in oleoresin flow in loblolly pine is controlled by many genes with allelic effects that are dependent on variation in climate and soils, we showed that it is feasible to develop genomic selection models that can accurately predict genetic variation in diverse environments. Work is underway to verify the significant oleoresin flow associations through analysis of gene expression, association genetic analysis in an independent population, and genetic transformation.

Acknowledgements

The Forest Biology Research Cooperative managed the CCLONES population and funded oleoresin collection. SNP genotyping was funded by the National Science Foundation Plant Genome Research Program award no. 0501763. Special thanks are due to L.H. Lott and J.H. Roberds (Southern Institute of Forest Genetics) for oleoresin taps; members of the University of Florida Forest Genomics Laboratory, especially J. Zhang, for assistance with oleoresin collection; M. Chavez Monte-Alegre for weighing the oleoresin tubes; and C. Dervinis and K. Smith for assistance with the BLAST analyses. J.W.W. was supported by a United States Department of Agriculture CSREES Food and Agricultural Sciences National Needs Graduate Fellowship.

References

- Aharoni A, Jongsma MA, Kim TY, Ri MB, Giri AP, Verstappen FWA, Schwab W, Bouwmeester HJ. 2006. Metabolic engineering of terpenoid biosynthesis in plants. *Phytochemistry Reviews* 5: 49–58.
- Alonso JM, Hirayama T, Roman G, Nourizadeh S, Ecker JR. 1999. EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. *Science* 284: 2148–2152.
- Aspinwall MJ, King JS, Booker FL, McKeand SE. 2011. Genetic effects on total phenolics, condensed tannins and non-structural carbohydrates in loblolly pine (*Pinus taeda* L.) needles. *Tree Physiology* 31: 831–842.
- Baltunis BS, Huber DA, White TL, Goldfarb B, Stelzer HE. 2007. Genetic analysis of early field growth of loblolly pine clones and seedlings from the same full-sib families. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* 37: 195–205.
- Bentz BJ, Regniere J, Fettig CJ, Hansen EM, Hayes JL, Hicke JA, Kelsey RG, Negron JF, Seybold SJ. 2010. Climate change and bark beetles of the western United States and Canada: direct and indirect effects. *BioScience* 60: 602–613.
- Bohlmann J, Meyer-Gauen G, Croteau R. 1998. Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proceedings of the National Academy of Sciences, USA* 95: 4126–4133.
- Boone CK, Aukema BH, Bohlmann J, Carroll AL, Raffa KF. 2011. Efficacy of tree defense physiology varies with bark beetle population density: a basis for positive feedback in eruptive species. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* 41: 1174–1188.

- Box GEP, Cox DR. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B* 26: 211–252.
- Bridgen MR. 1980. Genetic variation of Scotch pine oleoresin physiology. *Dissertation Abstracts International*, B 40: 4047B.
- Calder WJ, Horn KJ, St Clair SB. 2011. Conifer expansion reduces the competitive ability and herbivore defense of aspen by modifying light environment and soil chemistry. *Tree Physiology* 31: 582–591.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzalez-Martinez SC, Neale DB. 2010. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185: 969–982.
- Eckert AJ, Wegrzyn JL, Cumbie WP, Goldfarb B, Huber DA, Tolstikov V, Fiehn O, Neale DB. 2012. Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. *New Phytologist* 193: 890–902.
- Evans JM, Cohen MJ. 2009. Regional water resource implications of bioethanol production in the southeastern United States. *Global Change Biology* 15: 2261–2273.
- Franceschi VR, Krokene P, Christiansen E, Krekling T. 2005. Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytologist* 167: 353–375.
- Garrick DJ, Taylor JF, Fernando RL. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution* 41: 1–8.
- Gezan SA, White TL, Huber DA. 2006. Comparison of experimental designs for clonal forestry using simulated data. *Forest Science* 52: 108–116.
- Gillespie JH, Turelli M. 1989. Genotype–environment interactions and the maintenance of polygenic variation. *Genetics* 121: 129–138.
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R. 2009. *ASReml user guide release 3.0*. Hemel Hempstead, UK: VSN International Ltd.
- Gopal V, Li Z, Casella G. 2011. *BAMD: Bayesian association model for genomic data with missing covariates*. [WWW document] URL <http://cran.r-project.org/web/packages/BAMD/index.html> [accessed on 31 December 2011].
- Grattapaglia D, Resende MDV. 2011. Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7: 241–255.
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12: 186.
- Hamberger B, Ohnishi T, Hamberger B, Séguin A, Bohlmann J. 2011. Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiology* 157: 1677–1695.
- Harvey BG, Wright ME, Quintana RL. 2010. High-density renewable fuels based on the selective dimerization of pinenes. *Energy & Fuels* 24: 267–273.
- Hudgins JW, Franceschi VR. 2004. Methyl jasmonate-induced ethylene production is responsible for conifer phloem defense responses and reprogramming of stem cambial zone for traumatic resin duct formation. *Plant Physiology* 135: 2134–2149.
- Jasinski M, Stukkens Y, Degand H, Purnelle B, Marchand-Brynaert J, Boutry M. 2001. A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion. *Plant Cell* 13: 1095–1107.
- Jennewein S, Rithner CD, Williams RM, Croteau RB. 2000. Taxol biosynthesis: taxane 13 α -hydroxylase is a cytochrome P450-dependent monooxygenase. *Proceedings of the National Academy of Sciences, USA* 98: 13595–13600.
- Jirsitzka J, Mattern DJ, Gershenzon J, D'Auria JC. 2012. Learning from nature: new approaches to the metabolic engineering of plant defense pathways. *Current Opinion in Biotechnology* 24: 1–9.
- Kane JM, Kolb TE. 2010. Importance of resin ducts in reducing ponderosa pine mortality from bark beetle attack. *Oecologia* 164: 601–609.
- Keeling CI, Weisshaar S, Ralph SG, Jancsik S, Hamberger B, Dullat HK, Bohlmann J. 2011. Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). *BMC Plant Biology* 11: 1–14.
- Kenward MG, Roger JH. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53: 983–997.
- Kim GT, Fujioka S, Kozuka T, Tax FE, Takatsuto S, Yoshida S, Tsukaya H. 2005. CYP90C1 and CYP90D1 are involved in different steps in the brassinosteroid biosynthesis pathway in *Arabidopsis thaliana*. *Plant Journal* 41: 710–721.
- Kim YB, Kim SM, Kang MK, Kuzuyama T, Lee JK, Park SC, Shin SC, Kim SU. 2009. Regulation of resin acid synthesis in *Pinus densiflora* by differential transcription of genes encoding multiple 1-deoxy-d-xylulose 5-phosphate synthase and 1-hydroxy-2-methyl-2-(e)-butenyl 4-diphosphate reductase genes. *Tree Physiology* 29: 737–749.
- Kim SM, Kuzuyama T, Kobayashi A, Sando T, Chang YJ, Kim SU. 2008. 1-hydroxy-2-methyl-2-(e)-butenyl 4-diphosphate reductase (IDS) is encoded by multicopy genes in gymnosperms *Ginkgo biloba* and *Pinus taeda*. *Planta* 227: 287–298.
- Kopper BJ, Illman BL, Kersten PJ, Klepzig KD, Raffa KF. 2005. Effects of diterpene acids on components of a conifer bark beetle–fungal interaction: tolerance by *Ips pini* and sensitivity by its associate *Ophiostoma ips*. *Environmental Entomology* 34: 486–493.
- Kurz WA, Dymond CC, Stinson G, Rampley GJ, Neilson ET, Carroll AL, Ebata T, Safranyik L. 2008. Mountain pine beetle and forest carbon feedback to climate change. *Nature* 452: 987–990.
- Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, Tidor B, Stephanopoulos G, Prather KLJ. 2010. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proceedings of the National Academy of Sciences, USA* 107: 13654–13659.
- Li Z, Gopal V, Li X, Davis JM, Casella G. 2012. Simultaneous SNP identification in association studies with missing data. *Annals of Applied Statistics* 6: 432–456.
- Lorio PL. 1986. Growth-differentiation balance – a basis for understanding southern pine–beetle tree interactions. *Forest Ecology and Management* 14: 259–273.
- Lorio PL, Sommers RA. 1986. Evidence of competition for photosynthates between growth processes and oleoresin synthesis in *Pinus taeda* L. *Tree Physiology* 2: 301–306.
- van Mantgem PJ, Stephenson NL, Byrne JC, Daniels LD, Franklin JF, Fule PZ, Harmon ME, Larson AJ, Smith JM, Taylor AH *et al.* 2009. Widespread increase of tree mortality rates in the western United States. *Science* 323: 521–524.
- Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Moreno JE, Tao Y, Chory J, Ballare CL. 2009. Ecological modulation of plant defense via phytochrome control of jasmonate sensitivity. *Proceedings of the National Academy of Sciences, USA* 106: 4935–4940.
- Munoz PR, Resende MFR, Huber DA, Quesada T, Resende MDV, Neale DB, Wegrzyn JL, Kirst M, Peter GF. 2013. Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy. *Crop Science*, in press.
- Nicotra AB, Atkin OK, Bonser SP, Davidson AM, Finnegan EJ, Mathesius U, Poot P, Purugganan MD, Richards CL, Valladares F *et al.* 2010. Plant phenotypic plasticity in a changing climate. *Trends in Plant Science* 15: 684–692.
- Nowak J, Asaro C, Klepzig K, Billings R. 2008. The southern pine beetle prevention initiative: working for healthier forests. *Journal of Forestry* 106: 261–267.
- Ogawa T, Uchimiya H, Kawai-Yamada M. 2007. Mutual regulation of *Arabidopsis thaliana* ethylene-responsive element binding protein and a plant floral homeotic gene, APETALA2. *Annals of Botany* 99: 239–244.
- Paine TD, Raffa KF, Harrington TC. 1997. Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annual Review of Entomology* 42: 179–206.
- Patterson HD, Williams ER. 1976. New class of resolvable incomplete block designs. *Biometrika* 63: 83–92.
- Peralta-Yahya PP, Keasling JD. 2010. Advanced biofuel production in microbes. *Biotechnology Journal* 5: 147–162.
- Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS. 2011. Identification and microbial production of a terpene-based advanced biofuel. *Nature Communications* 2: 1–8.
- Phillips MA, Croteau RB. 1999. Resin-based defenses in conifers. *Trends in Plant Science* 4: 184–190.

- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. 2000. Association mapping in structured populations. *American Journal of Human Genetics* 67: 170–181.
- R Core Team. 2012. *R: a language and environment for statistical computing*. Vienna, Austria. [WWW document] URL <http://www.R-project.org/> [accessed on 31 December 2011].
- Raffa KF, Aukema BH, Bentz BJ, Carroll AL, Hicke JA, Turner MG, Romme WH. 2008. Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. *BioScience* 58: 501–517.
- Resende MFR, Munoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MDV, Kirst M. 2012a. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193: 617–624.
- Resende MFR, Munoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M. 2012b. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190: 1503–1510.
- Ro DK, Arimura GL, Lau SYW, Piers E, Bohlmann J. 2005. Loblolly pine abietadienol/abietadienol oxidase ptao (cyp720b1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. *Proceedings of the National Academy of Sciences, USA* 102: 8060–8065.
- Roberts JH, Strom BL, Hain FP, Gwaze DP, McKeand SE, Lott LH. 2003. Estimates of genetic parameters for oleoresin and growth traits in juvenile loblolly pine. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* 33: 2469–2476.
- Rodriguez-Concepcion M, Boronat A. 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiology* 130: 1079–1089.
- Romanelli RC, Sebbenn AM. 2004. Genetic parameters and selection gains for oleoresin production in *Pinus elliottii* var. *elliottii*, in south of São Paulo state. *Revista do Instituto Florestal* 16: 11–23.
- Schmidt A, Gershenzon J. 2008. Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*). *Phytochemistry* 69: 49–57.
- Schmidt A, Nagel R, Kreckling T, Christiansen E, Gershenzon J, Krokene P. 2011. Induction of isoprenyl diphosphate synthases, plant hormones and defense signalling genes correlates with traumatic resin duct formation in Norway spruce (*Picea abies*). *Plant Molecular Biology* 77: 577–590.
- Schulz-Streeck T, Ogutu JO, Piepho HP. 2011. Preselection of markers for genomic selection. *BMC Proceedings* 5(Suppl 3): S12.
- Squillace AE, Bengtson GW. 1961. Inheritance of gum yield and other characteristics of slash pine. In: *Proceedings of the Sixth Southern Conference on Forest Tree Improvement*. Gainesville, FL: School of Forestry, University of Florida, 85–96.
- Strom BL, Goyer RA, Ingram LL, Boyd GDL, Lott LH. 2002. Oleoresin characteristics of progeny of loblolly pines that escaped attack by southern pine beetle. *Forest Ecology and Management* 158: 169–178.
- Stubbs J, Roberts DR, Outcalt KW. 1984. Chemical stimulation of lightwood in Southern pines. In: *General Technical Report SE-25*, Asheville, NC, USA: US Forest Service, 1–50.
- Telewski FW, Wakefield AH, Jaffe MJ. 1983. Computer-assisted image-analysis of tissues of ethrel-treated *Pinus taeda* seedlings. *Plant Physiology* 72: 177–181.
- Trapp S, Croteau R. 2001. Defensive resin biosynthesis in conifers. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 689–724.
- Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG. 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLOS Genetics* 2: 316–325.
- Wallin KF, Kolb TE, Skov KR, Wagner M. 2008. Forest management treatments, tree resistance, and bark beetle resource utilization in ponderosa pine forests of northern Arizona. *Forest Ecology and Management* 255: 3263–3269.
- Wang Y, Lim L, Robertson G, Bohlmann J, Breuil C. 2013. A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grossmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. *New Phytologist* 197: 886–898.
- Waring KM, Reboletti DM, Mork LA, Huang CH, Hofstetter RW, Garcia AM, Fule PZ, Davis TS. 2009. Modeling the impacts of two bark beetle species under a warming climate in the southwestern USA: ecological and economic consequences. *Environmental Management* 44: 824–835.
- White TL, Carson M. 2004. Breeding program of conifers. In: Walter C, Carson M, eds. *Plantation forest biotechnology for the 21st century*. Research Signpost, 61–85.
- Wolter KE, Zinkel DF. 1984. Observations on the physiological-mechanisms and chemical-constituents of induced oleoresin synthesis in *Pinus resinosa*. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* 14: 452–458.
- Yang JA, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW *et al.* 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42: 565–569.
- Zeller SL, Kalinina O, Brunner S, Keller B, Schmid B. 2010. Transgene x environment interactions in genetically modified wheat. *PLoS ONE* 5: 7.
- Zulak KG, Bohlmann J. 2010. Terpenoid biosynthesis and specialized vascular cells of conifer defense. *Journal of Integrative Plant Biology* 52: 86–97.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Oleoresin wet mass and dry mass comparisons.

Fig. S2 A comparison of SNP preselection methods with cross-validation.

Table S1 Formulas for the estimation of heritabilities and correlations

Table S2 Contrasts of average transformed oleoresin dry mass between sites and years

Table S3 Summary of significant SNP associations with transformed oleoresin dry mass

Table S4 Site-specific associations with transformed oleoresin dry mass

Table S5 Tests for associations with SNPs within sequences similar to terpene biosynthetic genes

Table S6 Predicted F1 gains from selection for increased oleoresin flow

Methods S1 Oleoresin dry mass transformation.

Methods S2 Likelihood ratio tests for correlations.

Methods S3 Preselection of SNP loci for association testing.

Methods S4 Tests for subpopulation structure.

Methods S5 Predicting gains from selection.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.